ExerciseScargle: Introduction

ExerciseSineFit is an example of how to fit a sinusoidal model **G** to the data t_i = **T**, y_i = **Y** and σ_i = **EY** **when** the period P = **P** is **known**. The current **ExerciseScargle** shows what can be done, if the period P = **P** is **unknown?**

If the data contains a sinusoidal signal, the **Discrete Fourier Transform** (DFT) can detect the period $P_{\rm best}$ = **PBEST** of this signal. This method is also called the **Power Spectrum Method**. Here, we use the formulation by [1]Scargle (1982, his Eqs. 10 and 11). In August 2020, this paper had over four thousand citations in ADS. This means that DFT is probably the most frequently applied period finding method in Astronomy. However, DFT is a **one-dimensional period finding method**, because it can detect only one period at the time. DFT periodogram is

$$z_{\rm LS}(f) = \frac{1}{2}\left[\frac{[\Sigma_{\rm i=1}^{\rm n} y'_{\rm i}\cos 2\pi f(t_{\rm i}-\tau)]^2}{\Sigma_{\rm i=1}^{\rm n}[\cos 2\pi f(t_{\rm i}-\tau)]^2} + \frac{[\Sigma_{\rm i=1}^{\rm n} y'_{\rm i}\sin 2\pi f(t_{\rm i}-\tau)]^2}{\Sigma_{\rm i=1}^{\rm n}[\sin 2\pi f(t_{\rm i}-\tau)]^2}\right] = \frac{1}{2}\left[\frac{z_1}{z_2}+\frac{z_3}{z_4}\right]$$

$$\tau = [1/(4\pi f)]\ {\rm atan}\left[\left(\sum_{\rm i=1}^{\rm n}\sin 4\pi f t_{\rm i}\right) / \left(\sum_{\rm i=1}^{\rm n}\cos 4\pi f t_{\rm i}\right)\right]$$

where z_{LS} = **ZLS** and τ = **TAU** are computed from
t_i = **T** = time points
$y'_i = y_i - m_{\rm y}$ =**YDOT=Y-np.mean(Y)**
m_y = **np.mean(Y)** = mean of y_i
f = **F** = tested frequency

Note: DFT ignores σ_i error information!

Our notations for the sums are
$z_1 = [\Sigma_{\mathrm{i}=1}^{\mathrm{n}}\, y'_{\mathrm{i}} \cos 2\pi f(t_{\mathrm{i}} - \tau)]^2 =$ **Z1**
$z_2 = \Sigma_{\mathrm{i}=1}^{\mathrm{n}}[\cos 2\pi f(t_{\mathrm{i}} - \tau)]^2 =$ **Z2**
$z_3 = [\Sigma_{\mathrm{i}=1}^{\mathrm{n}}\, y'_{\mathrm{i}} \sin 2\pi f(t_{\mathrm{i}} - \tau)]^2 =$ **Z3**
$z_4 = \Sigma_{\mathrm{i}=1}^{\mathrm{n}}[\sin 2\pi f(t_{\mathrm{i}} - \tau)]^2 =$ **Z4**

- The tested period range is between $P_{\mathrm{min}} =$ **PMIN** and $P_{\mathrm{max}} =$ **PMAX**. This gives the tested frequency range between $1/P_{\mathrm{max}} = f_{\mathrm{min}} =$ **FMIN** and $1/P_{\mathrm{min}} = f_{\mathrm{max}} =$ **FMAX**.

- The step in tested freqeuncies is $f_{\mathrm{step}} = f_0/(\mathrm{OFAC}) =$ **FSTEP**, where $f_0 = 1/\Delta T$ and $\Delta T = t_n - t_1 =$ **np.max(T)-np.min(T)=DELTAT** is the time span of data, and $\mathrm{OFAC} =$ **OFAC** is the over filling factor.

- The tested frequencies $f =$ **F** are all integer multiples of $f_{\mathrm{step}} =$ **FSTEP** between $f_{\mathrm{min}} =$ **FMIN** and $f_{\mathrm{max}} =$ **FMAX**

Main point: For a purely sinusoidal model, the best frequency $f_{\mathrm{best}} =$ **FBEST** is at the DFT periodogram maximum $z_{\mathrm{LS,max}}(f_{\mathrm{best}}) =$ **ZBEST**. The best period $P_{\mathrm{best}} = 1/f_{\mathrm{best}} =$ **1./FBEST**.

ExerciseScargle: Problem

1. Download file **Scargle.dat** from homepage.
2. Give the name **ExerciseScargle.py** to your **python** solution program.
3. Read data **T**, **Y** and **EY** from **Scargle.dat**. Note that the errors **EY** are not needed in this exercise.
4. Compute and plot the DFT periodogram for these data.
5. Use same variable names that are used in this document, like $t_i =$ **T**.
6. Fix input parameters to **PMIN=1.0**, **PMAX=5.0** and **OFAC=40**.

7. Compute DFT periodogram $z_{\rm LS}(f)$ = **ZLS** with a subroutine **F,ZLS,FBEST,ZBEST=periodogram(PMIN,PMAX,OFAC,T,Y)**, where the highest **ZLS** peak is at frequency **FBEST**, and the height of this peak is **ZBEST**.
8. Plot your results for the periodogram as in **figure** on the last page. Indicate the highest peak with a circle. Note that the **content** should be the same as in this **figure**, but not the **form** (e.g. symbols, panels).
9. Give the numerical best period $P_{\rm best}$ = **1/FBEST** above your plot.
10. Store your figure/plot to the file **ExerciseScargle.eps**.
11. Send **ExerciseScargle.py** and **ExerciseScargle.eps** to the assistant.

Some tips

- Keep input, computations and plots apart in different subroutines.
- Here is one example of how to read data

```
def Readfile(file2):
    T=np.loadtxt(file2,skiprows=0,usecols=(0,))
    Y=np.loadtxt(file2,skiprows=0,usecols=(1,))
    EY=np.loadtxt(file2,skiprows=0,usecols=(2,))
    return T,Y,EY
```

- Number of tested frequencies is **m=np.floor((FMAX-FMIN)/FSTEP)**
- Create frequency vector **F** first. Then create periodogram vector **ZLS=0.*F**
- **Begin a frequency F[i] loop**.
  - **For each tested frequency F[i]**:
  - **First** compute τ =**TAU**
  - **Then** compute $z_{LS}(f)$ =**ZLS[i]**
- Simplify your program,
  like using argument $2\pi f(t_i - \tau)$ =**x2=2.*np.pi*F[i]*(T-TAU)**,
  which gives z_1 = **z1=(np.sum(ydot*np.cos(x2)))**2.**

- After frequency **F[i]** loop, the command
  j=(ZLS==np.max(ZLS)).nonzero()
  gives index of **ZLS** maximum.

References

[1] J. D. Scargle. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, 263:835–853, Dec. 1982.